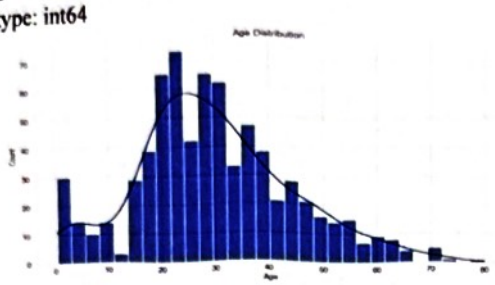
2 Exploratory data analysis (EDA) visualizing datasets using matplotlib and Seaborn Identify trends, outliers and correlations

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_style('white grid')
plt.rcparams['figure.figsize'] = (10,6)
df = sns.load_dataset('titanic')
print("dataset Head")
print( df.head())
print("\n dataset info ;")
print ( df.info())
print ("\n Summary statistics")
print (df.describe())
print ("\n missing values")
print (df.isnull().sum())
plt.figure()
sns.histplot (df['age'].bins = 30 , kde = True , color = 'blue')
plt.title ('Age Distribution')
plt.xlabel ('age')
plt.ylabel ('count')
plt.show()
plt.figure()
```

```
sns.countplot(x='class', data=df, hallete="set2")
plt.title('Passenger class Distribution')
plt.xlabel('Passenger class')
plt.ylabel('count')
plt.show()
plt.figure()
sns.boxplot(x='class', y='age', data=df, hallete='set3')
Plt.title('Age distribution by passenger class')
plt.show()
plt.figure()
sns.lineplot(x='age', y='survived', data=df)
plt.title('survival trend by age')
plt.xlabel('age')
plt.ylabel('survival rate')
plt.show()
numerical.cols = df.select_dtypes(include=['float64','int64']).
  columns
coll = df[numerical.cols].coll()
plt.figure()
sns.heatmap(coll, annot=true, Cmap='coolwarm', Center=0)
plt.title('correlation Heat map')
plt.show()
plt.figure()
sns.scatterplot(x='age', y='fare', hue='survived', data=df.
  hallete='deep')
plt.title('rare vs age (coloured by survived)') plt.show()
```

```
Dataset Head:
  survived pclass    sex  age sibsp parch    fare embarked  class \
0        0      3   male 22.0     1     0  7.2500        S  Third
1        1      1 female 38.0     1     0 71.2833        C  First  2        1      3 female 26.0      0
         0  7.9250        S  Third
3        1      1 female 35.0     1     0 53.1000        S  First
4        0      3   male 35.0     0     0  8.0500        S  Third

    who  adult_male deck  embark_town alive  alone  0
man       True  NaN Southampton    no  False
1  woman     False    C    Cherbourg   yes  False
2  woman     False  NaN Southampton   yes   True
3  woman     False    C Southampton   yes  False
4    man      True  NaN Southampton    no   True

Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890 Data
columns (total 15 columns):
 #  Column      Non-Null Count  Dtype
---  ------      --------------  -----
0   survived    891 non-null    int64
1   pclass      891 non-null    int64
2   sex         891 non-null    object
3   age         714 non-null    float64
4   sibsp       891 non-null    int64
5   parch       891 non-null    int64
6   fare        891 non-null    float64
7   embarked    889 non-null    object
8   class       891 non-null    category
9   who         891 non-null    object
10  adult_male  891 non-null    bool
11  deck        203 non-null    category
12  embark_town 889 non-null    object
13  alive       891 non-null    object  14  alone       891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
None

Summary Statistics:
       survived     pclass       age     sibsp     parch      fare
count  891.000000  891.000000  714.000000  891.000000  891.000000  891.000000 mean
0.383838    2.308642   29.699118    0.523008    0.381594   32.204208 std       0.486592
0.836071   14.526497    1.102743    0.806057   49.693429 min       0.000000    1.000000
0.420000    0.000000    0.000000    0.000000 25%       0.000000    2.000000   20.125000
0.000000    0.000000    7.910400
50%     0.000000    3.000000   28.000000    0.000000    0.000000   14.454200 75%
1.000000    3.000000   38.000000    1.000000    0.000000   31.000000 max      1.000000
3.000000   80.000000    8.000000    6.000000  512.329200

Missing Values:
survived       0
pclass         0 sex
0 age          177
sibsp          0
parch          0
fare               0
embarked           2
```

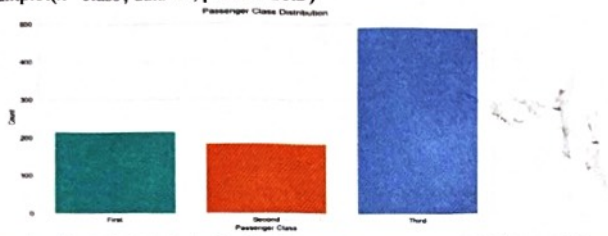
```
class          0 who
0 adult_male      0
deck          688
embark_town     2
alive          0 alone
0 dtype: int64
```



/var/folders/w_/tlx3kfln2s9btjy8l26bm8dw0000gn/T/ipykernel_1308/3123855039.py:71: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable t o `hue` and set `legend=False` for the same effect.

```
sns.countplot(x='class', data=df, palette='Set2')
```



/var/folders/w_/tlx3kfln2s9btjy8l26bm8dw0000gn/T/ipykernel_1308/3123855039.py:89: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable t o `hue` and set `legend=False` for the same effect.

```
sns.boxplot(x='class', y='age', data=df, palette='Set3')
```